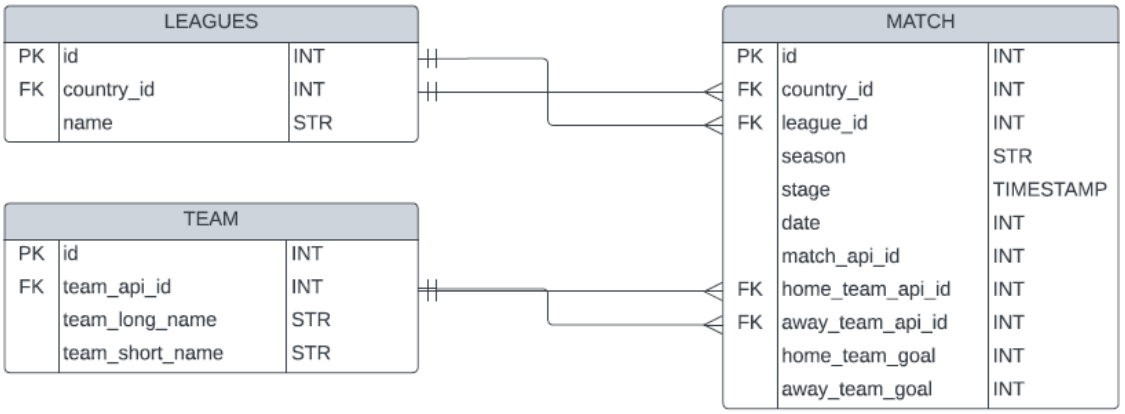
DATA EXPLORATION - EUROPEAN SOCCER DATABASE



---

1. How many days have passed from the oldest **Match** to the most recent one (dataset time interval)?

ANSWER: 2868 days

QUERY:

```
SELECT
        MAX(date) AS Newest_date,
        MIN(date) AS Oldest_date,
        TIMESTAMP_DIFF(MAX(date),MIN(date), DAY) AS Days_Dataset_inteval
FROM `progetto-di-prova-365418.Final_Exercise.Match`
```

---

2. Produce a table which, for each Season and **League** Name, shows some statistics (min, avg, mid-range, max, sum) about the home goals scored. Which combination of Season-League has the highest number of home goals?

ANSWER: England Premier League - Season 2009/2010 with 645 home goals.

QUERY:

```
SELECT match.season,
      leagues.name AS league_name,
      MIN(match.home_team_goal) AS min_homegoals,
      MAX(match.home_team_goal) AS max_homegoals,
      ROUND(AVG(match.home_team_goal),2) AS avg_homegoals,
      (MIN(match.home_team_goal) + MAX(match.home_team_goal))/2 AS midrange_homegoals,
      SUM(match.home_team_goal) AS sum_homegoals,
FROM `progetto-di-prova-365418.Final_Exercise.Match` AS match
LEFT JOIN `progetto-di-prova-365418.Final_Exercise.Leagues` AS leagues
on match.league_id = leagues.id
GROUP BY match.season, league_name
ORDER BY sum_homegoals DESC
LIMIT 1;
```

---

3. Find out how many unique seasons there are in the **Match** table.
Then write a query that shows, for each Season, the number of matches played by each League. Do you notice anything out of the ordinary?
ANSWER: The number of seasons in the "match" table is 8. It should be noted that in the 2013/2014 season only 12 matches were played in the Belgium Jupiler League.

QUERY:

```
SELECT COUNT(DISTINCT season) AS distinct_season
FROM `progetto-di-prova-365418.Final_Exercise.Match`;

SELECT
      match.season,
      leagues.name AS league_name,
      COUNT(match.id) AS number_matches,
FROM `progetto-di-prova-365418.Final_Exercise.Match` AS match
LEFT JOIN `progetto-di-prova-365418.Final_Exercise.Leagues` AS leagues
on match.league_id = leagues.id
GROUP BY match.season, league_name
ORDER BY number_matches;
```

---

4. Using Players as the starting point, create a new table (PlayerBMI) and add:
a. a new variable that represents the players' weight in kg (divide the mass value by 2.205) and call it kg_weight;
b. a variable that represents the height in metres (divide the cm value by 100) and call it m_height;
c. a variable that shows the body mass index (BMI) of the player;
d. Filter the table to show only the players with an optimal BMI (from 18.5 to 24.9).
How many rows does this table have? How many players do not have an optimal BMI?

ANSWER: Players with an excellent BMI are 10197. Players who do not have an excellent BMI are 863.

QUERY:

```
CREATE TABLE `progetto-di-prova-365418.Final_Exercise.PlayerBMI` AS
SELECT *,
      weight/2.205 AS kg_weight,
      height/100 AS m_height,
      (weight/2.205) / ((height/100)*(height/100)) AS BMI
FROM `progetto-di-prova-365418.Final_Exercise.Player`;

SELECT COUNT(*) AS N_player_with_optimal_BMI,
FROM `progetto-di-prova-365418.Final_Exercise.PlayerBMI`
WHERE BMI >= 18.5 AND BMI <= 24.9;

SELECT COUNT(*) AS N_player_with_not_optimal_BMI,
FROM `progetto-di-prova-365418.Final_Exercise.PlayerBMI`
WHERE BMI < 18.5 OR BMI > 24.9;
```

---

5. Which **Team** has scored the highest total number of goals (home + away) during the most recent available season? How many goals has it scored?

ANSWER: The team that has scored the highest goals number goals in the most recent season (2015/16) is Barcelona with 112 goals.

QUERY:

```
SELECT MAX(DISTINCT season) AS season_most_recent
FROM `progetto-di-prova-365418.Final_Exercise.Match`;

SELECT league_name, team_name, SUM(team_goal) AS tot_team_goal
FROM(
    SELECT * FROM
      (SELECT
            leagues.name AS league_name,
            team_home.team_long_name AS team_name,
            SUM(match.home_team_goal) AS team_goal,
      FROM `progetto-di-prova-365418.Final_Exercise.Match` AS match
      LEFT JOIN `progetto-di-prova-365418.Final_Exercise.Leagues` AS leagues
      on match.league_id = leagues.id
      LEFT JOIN `progetto-di-prova-365418.Final_Exercise.Team` AS team_home
      on match.home_team_api_id = team_home.team_api_id
      WHERE season = "2015/2016"
      GROUP BY league_name, team_name
      ) AS table_home_goals
    UNION ALL
    SELECT * FROM
      (SELECT
            leagues.name AS league_name,
            team_away.team_long_name AS team_name,
            SUM(match.away_team_goal) AS team_goal,
      FROM `progetto-di-prova-365418.Final_Exercise.Match` AS match
      LEFT JOIN `progetto-di-prova-365418.Final_Exercise.Leagues` AS leagues
      on match.league_id = leagues.id
      LEFT JOIN `progetto-di-prova-365418.Final_Exercise.Team` AS team_away
      on match.away_team_api_id = team_away.team_api_id
      WHERE season = "2015/2016"
      GROUP BY league_name, team_name
      ) AS table_away_goals
    )
GROUP BY league_name, team_name
ORDER BY tot_team_goal DESC
LIMIT 1;
```

---

6. Create a query that, for each season, shows the name of the team that ranks first in terms of total goals scored. Which team was the one that ranked first in most of the seasons?

ANSWER: The team that comes first most times for goals scored in the seasons played (among those available in the Dataset) is "Real Madrid CF" (4 times out of 8 seasons), followed by "FC Barcelona" (3 times out of 8 seasons) and “Ajax” (1 time in 8 seasons).

QUERY:

```
SELECT team_name, count(team_name) as num_seasons_as_1stForGoalScored,
FROM(
  SELECT *,
        RANK() over(partition by season order by tot_team_goal DESC) AS Frist_team_per_goal_each_season
  FROM(
      SELECT season, league_name, team_name, SUM(team_goal) AS tot_team_goal,
      FROM(
          SELECT * FROM
            (SELECT
                  match.season,
                  leagues.name AS league_name,
                  team_home.team_long_name AS team_name,
                  SUM(match.home_team_goal) AS team_goal,
            FROM `progetto-di-prova-365418.Final_Exercise.Match` AS match
            LEFT JOIN `progetto-di-prova-365418.Final_Exercise.Leagues` AS leagues
            on match.league_id = leagues.id
            LEFT JOIN `progetto-di-prova-365418.Final_Exercise.Team` AS team_home
            on match.home_team_api_id = team_home.team_api_id
            GROUP BY match.season, league_name, team_name
            ) AS table_home_goals
          UNION ALL
          SELECT * FROM
            (SELECT
                  match.season,
                  leagues.name AS league_name,
                  team_away.team_long_name AS team_name,
                  SUM(match.away_team_goal) AS team_goal,
            FROM `progetto-di-prova-365418.Final_Exercise.Match` AS match
            LEFT JOIN `progetto-di-prova-365418.Final_Exercise.Leagues` AS leagues
            on match.league_id = leagues.id
            LEFT JOIN `progetto-di-prova-365418.Final_Exercise.Team` AS team_away
            on match.away_team_api_id = team_away.team_api_id
            GROUP BY match.season, league_name, team_name
            ) AS table_away_goals
          )
      GROUP BY season, league_name, team_name
      )
    )
WHERE Frist_team_per_goal_each_season = 1
GROUP BY team_name;
```

---

7. From the query above create a new table (TopScorer) containing the top 10 teams in terms of total goals scored. Then write a query that shows all the possible "pair combinations" between those 10 teams. How many "pair combinations" did it generate?

ANSWER: The top 10 teams for goals scored in the "TopScorer" table are: FC Barcelona, Real Madrid CF, Celtic, FC Bayern Munich, PSV, Ajax, FC Basel, Manchester City, Chelsea, Manchester United.
Considering all matches in which the "TopScorer" teams face each other in each championship, there are a total of 80 matches (10 for each season among those available in the dataset).

QUERY:

```
CREATE TABLE `progetto-di-prova-365418.Final_Exercise.TopScorer` AS
SELECT team_name, SUM(team_goal) AS tot_team_goal,
FROM(
    SELECT * FROM
      (SELECT
            match.season,
            leagues.name AS league_name,
            team_home.team_long_name AS team_name,
            SUM(match.home_team_goal) AS team_goal,
      FROM `progetto-di-prova-365418.Final_Exercise.Match` AS match
      LEFT JOIN `progetto-di-prova-365418.Final_Exercise.Leagues` AS leagues
      on match.league_id = leagues.id
      LEFT JOIN `progetto-di-prova-365418.Final_Exercise.Team` AS team_home
      on match.home_team_api_id = team_home.team_api_id
      GROUP BY match.season, league_name, team_name
      ) AS table_home_goals
    UNION ALL
    SELECT * FROM
      (SELECT
            match.season,
            leagues.name AS league_name,
            team_away.team_long_name AS team_name,
            SUM(match.away_team_goal) AS team_goal,
      FROM `progetto-di-prova-365418.Final_Exercise.Match` AS match
      LEFT JOIN `progetto-di-prova-365418.Final_Exercise.Leagues` AS leagues
      on match.league_id = leagues.id
      LEFT JOIN `progetto-di-prova-365418.Final_Exercise.Team` AS team_away
      on match.away_team_api_id = team_away.team_api_id
      GROUP BY match.season, league_name, team_name
      ) AS table_away_goals
    )
GROUP BY team_name
ORDER BY tot_team_goal DESC
LIMIT 10;


SELECT season, COUNT(match_id) AS NumTopMatches
FROM  (
        SELECT  id as match_id,
                season,
                home_team_api_id,
                away_team_api_id,
                team_long_name_home,
                team_long_name_away,
```

```
              CASE WHEN team_long_name_home IN (SELECT team_name FROM
`progetto-di-prova-365418.Final_Exercise.TopScorer`) THEN "TopScorer"
                ELSE "NormalScorer"
                END AS ScorerType_home,
              CASE WHEN team_long_name_away IN (SELECT team_name FROM
`progetto-di-prova-365418.Final_Exercise.TopScorer`) THEN "TopScorer"
                ELSE "NormalScorer"
                END AS ScorerType_away
       FROM  (
              SELECT match.*, team.team_long_name as team_long_name_away
              FROM(
                    SELECT match.*, team.team_long_name as team_long_name_home
                    FROM `progetto-di-prova-365418.Final_Exercise.Match` AS match
                    LEFT JOIN `progetto-di-prova-365418.Final_Exercise.Team` AS team
                      on  match.home_team_api_id = team.team_api_id
                    ) as match
              LEFT JOIN `progetto-di-prova-365418.Final_Exercise.Team` AS team
                on  match.away_team_api_id = team.team_api_id
            )
     )
WHERE ScorerType_home = "TopScorer" AND ScorerType_away = "TopScorer"
GROUP BY season
```
